

Group44 project Life Expectancy

null

Kaiyan Zhu_B00797725,Jiaxing Huang_B00753848,Xinyu(Louis) Lu_B00799504,sun yu_B00792609

2020/12/11

- Abstract
- Introduction
- Data Description
 - Data loading and Required libraries
 - Data Description
 - General layout
 - Summary Statistics
 - Missing values
 - Correlation Analysis and Correlation Plots
 - Data Visualization
 - Histograms
 - Box plot
 - Scatter plots
- Methods
 - Regression
 - Analysis of Variance
 - Stepwise selection
 - Summary Statistics and ANOVA
- Results
 - Model Residuals
- Conclusion
- Appendix

Abstract

The level of life expectancy varies across the countries and with that, the variation in the expenditure on health, influenced by the level of Gross Domestic Product attained by a country. The health care provided from birth of an infant, in terms of the immunization and vaccination at the early age determine the mortality rate. The health care provided to the entire population determines the mortality rate. The analysis of the demographic, socio-economic, immunization and mortality rates within countries to assess their influence on the life expectancy is important for policy making process.

Introduction

An analysis of the life expectancy, over periods of time across the developed and developing countries is undertaken herein. The focus is on how the other variables within the data provided relate with life expectancy. Trends in the life expectancy across the years are undertaken based on the status of the countries and/or under any other suitable variable within our data. Visualization to assess the distribution of our variable of interest and how it relates to other variables are made and interpreted. The interpretations are further used to support our initial assumptions developed within the analysis stages. A regression model is fit and an analysis of variance on the model is made to assess the effectiveness of the variables in predicting the life expectancy. Step analysis model is made to assess the significant variables and compared with other models to check for any significant difference.

Data Description

The **Life Expectancy (WHO)** data set was obtained from the Kaggle (<https://www.kaggle.com/kumarajarshi/life-expectancy-who>).

Data loading and Required libraries

```
library(tidyverse)
library(readr)

Life_Expectancy_Data <- read_csv("D:/STAT3340/project/Data.csv")
```

Data Description

General layout

The **Country**, **Year** and **Status** variables are converted to factors for ease of analysis due to their distinct nature. When the data is passed as an argument to `dplyr::glimpse()` function, we are able to see the general layout of the data attributes.

The data contains 2938 instances, 22 attributes.

```
dim(Life_Expectancy_Data)
```

```
## [1] 2938 22
```

```
colnames(Life_Expectancy_Data)
```

```
## [1] "Country"           "Year"
## [3] "Status"            "Life expectancy"
## [5] "Adult Mortality"   "infant deaths"
## [7] "Alcohol"           "percentage expenditure"
## [9] "Hepatitis B"       "Measles"
## [11] "BMI"               "under-five deaths"
## [13] "Polio"             "Total expenditure"
## [15] "Diphtheria"        "HIV/AIDS"
## [17] "GDP"               "Population"
## [19] "thinness 1-19 years" "thinness 5-9 years"
## [21] "Income composition of resources" "Schooling"
```

There are categorical attributes as well as numeric attributes within. The data covers 193 countries, with the country classified as either Developed or Developing under the **Status** variable. The period under which the data is considered is from the year 2001 through to 2015.

```
#Unique values
length(unique(Life_Expectancy_Data$Country))
```

```
## [1] 193
```

```
unique(Life_Expectancy_Data$Year)
```

```
## [1] 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001
## [16] 2000
```

```
unique(Life_Expectancy_Data$Status)
```

```
## [1] "Developing" "Developed"
```

```
Life_Expectancy_Data$Country <- as.factor(Life_Expectancy_Data$Country)
Life_Expectancy_Data$Year <- as.factor(Life_Expectancy_Data$Year)
Life_Expectancy_Data$Status <- as.factor(Life_Expectancy_Data$Status)
#Glimpse of Structure
glimpse(Life_Expectancy_Data)
```

```
## Rows: 2,938
## Columns: 22
## $ Country          <fct> Afghanistan, Afghanistan, Afghani...
## $ Year              <fct> 2015, 2014, 2013, 2012, 2011, 201...
## $ Status            <fct> Developing, Developing, Developin...
## $ `Life expectancy` <dbl> 65.0, 59.9, 59.9, 59.5, 59.2, 58...
## $ `Adult Mortality` <dbl> 263, 271, 268, 272, 275, 279, 281...
## $ `infant deaths`   <dbl> 62, 64, 66, 69, 71, 74, 77, 80, 8...
## $ Alcohol           <dbl> 0.01, 0.01, 0.01, 0.01, 0.01, 0.0...
## $ `percentage expenditure` <dbl> 71.279624, 73.523582, 73.219243, ...
## $ `Hepatitis B`    <dbl> 65, 62, 64, 67, 68, 66, 63, 64, 6...
## $ Measles           <dbl> 1154, 492, 430, 2787, 3013, 1989,...
## $ BMI               <dbl> 19.1, 18.6, 18.1, 17.6, 17.2, 16...
## $ `under-five deaths` <dbl> 83, 86, 89, 93, 97, 102, 106, 110...
## $ Polio             <dbl> 6, 58, 62, 67, 68, 66, 63, 64, 63...
## $ `Total expenditure` <dbl> 8.16, 8.18, 8.13, 8.52, 7.87, 9.2...
## $ Diphtheria        <dbl> 65, 62, 64, 67, 68, 66, 63, 64, 6...
## $ `HIV/AIDS`       <dbl> 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1...
## $ GDP               <dbl> 584.25921, 612.69651, 631.74498, ...
## $ Population        <dbl> 33736494, 327582, 31731688, 36969...
## $ `thinness 1-19 years` <dbl> 17.2, 17.5, 17.7, 17.9, 18.2, 18...
## $ `thinness 5-9 years` <dbl> 17.3, 17.5, 17.7, 18.0, 18.2, 18...
## $ `Income composition of resources` <dbl> 0.479, 0.476, 0.470, 0.463, 0.454...
## $ Schooling         <dbl> 10.1, 10.0, 9.9, 9.8, 9.5, 9.2, 8...
```

Summary Statistics

The summary statistic provides the count for categoric variables, the minimum, maximum, 1st quantile, median, mean, 3rd quantile and the number of missing data points for numeric variables.

```
options(scipen = 999)
summary(Life_Expectancy_Data[,c(-1,-2)])
```

```
##      Status      Life expectancy Adult Mortality infant deaths
## Developed : 512   Min.    :36.30   Min.    : 1.0   Min.    : 0.0
## Developing:2426  1st Qu.:63.10   1st Qu.: 74.0   1st Qu.: 0.0
##                Median :72.10   Median :144.0   Median : 3.0
##                Mean   :69.22   Mean    :164.8   Mean    : 30.3
##                3rd Qu.:75.70   3rd Qu.:228.0   3rd Qu.: 22.0
##                Max.    :89.00   Max.    :723.0   Max.    :1800.0
##                NA's    :10     NA's    :10
##      Alcohol      percentage expenditure Hepatitis B      Measles
## Min.    : 0.0100   Min.    : 0.000   Min.    : 1.00   Min.    : 0.0
## 1st Qu.: 0.8775   1st Qu.: 4.685   1st Qu.:77.00   1st Qu.: 0.0
## Median : 3.7550   Median : 64.913   Median :92.00   Median : 17.0
## Mean    : 4.6029   Mean    : 738.251   Mean    :80.94   Mean    : 2419.6
## 3rd Qu.: 7.7025   3rd Qu.: 441.534   3rd Qu.:97.00   3rd Qu.: 360.2
## Max.    :17.8700   Max.    :19479.912   Max.    :99.00   Max.    :212183.0
## NA's    :194     NA's    :553
##      BMI      under-five deaths      Polio      Total expenditure
## Min.    : 1.00   Min.    : 0.00   Min.    : 3.00   Min.    : 0.370
## 1st Qu.:19.30   1st Qu.: 0.00   1st Qu.:78.00   1st Qu.: 4.260
## Median :43.50   Median : 4.00   Median :93.00   Median : 5.755
## Mean    :38.32   Mean    : 42.04   Mean    :82.55   Mean    : 5.938
## 3rd Qu.:56.20   3rd Qu.: 28.00   3rd Qu.:97.00   3rd Qu.: 7.492
## Max.    :87.30   Max.    :2500.00   Max.    :99.00   Max.    :17.600
## NA's    :34     NA's    :19     NA's    :226
##      Diphtheria      HIV/AIDS      GDP      Population
## Min.    : 2.00   Min.    : 0.100   Min.    : 1.68   Min.    : 34
## 1st Qu.:78.00   1st Qu.: 0.100   1st Qu.: 463.94   1st Qu.: 195793
## Median :93.00   Median : 0.100   Median : 1766.95   Median : 1386542
## Mean    :82.32   Mean    : 1.742   Mean    : 7483.16   Mean    : 12753375
## 3rd Qu.:97.00   3rd Qu.: 0.800   3rd Qu.: 5910.81   3rd Qu.: 7420359
## Max.    :99.00   Max.    :50.600   Max.    :119172.74   Max.    :1293859294
## NA's    :19     NA's    :448   NA's    :652
##      thinness 1-19 years thinness 5-9 years Income composition of resources
## Min.    : 0.10   Min.    : 0.10   Min.    :0.0000
## 1st Qu.: 1.60   1st Qu.: 1.50   1st Qu.:0.4930
## Median : 3.30   Median : 3.30   Median :0.6770
## Mean    : 4.84   Mean    : 4.87   Mean    :0.6276
## 3rd Qu.: 7.20   3rd Qu.: 7.20   3rd Qu.:0.7790
## Max.    :27.70   Max.    :28.60   Max.    :0.9480
## NA's    :34     NA's    :34     NA's    :167
##      Schooling
## Min.    : 0.00
## 1st Qu.:10.10
## Median :12.30
## Mean    :11.99
## 3rd Qu.:14.30
## Max.    :20.70
## NA's    :163
```

The `Population` variable contains the highest number of missing values followed by the `Hepatitis B` variable.

Missing values

```
#Missing values
sum(is.na(Life_Expectancy_Data))
```

```
## [1] 2563
```

```
sum(is.na(Life_Expectancy_Data$`Life expectancy`))
```

```
## [1] 10
```

```
sum(is.na(Life_Expectancy_Data$Population))
```

```
## [1] 652
```

There are 2563 missing values in the whole data, with 10 being within the Life Expectancy attribute and 652 within the Population which we can say may be due to uncollected data.

Correlation Analysis and Correlation Plots

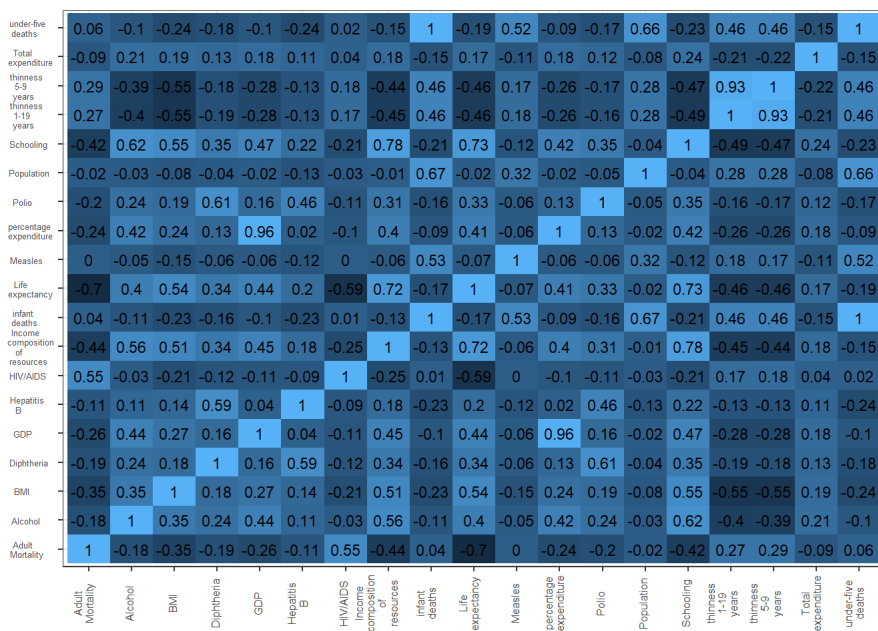
The linear relation among the variable in our data is found and the correlation table plotted.

```
#Correlation
library(ggplot2)
library(reshape) # to generate input for the plot

correlation_matrix <- round(
  cor(na.omit(Life_Expectancy_Data[,c(-1,-2,-3)])),2
) # rounded correlation matrix

melted_correlation_matrix <- melt(correlation_matrix)
melted_correlation_matrix$X1 <- as.factor(melted_correlation_matrix$X1)
melted_correlation_matrix$X2 <- as.factor(melted_correlation_matrix$X2)
#Remove the extra white spaces and existing replace with newline for
#axis text optimization
levels(melted_correlation_matrix$X1) <- gsub(
  " ", "\n", str_squish(
    levels(melted_correlation_matrix$X1)
  )
)
levels(melted_correlation_matrix$X2) <- gsub(
  " ", "\n", str_squish(
    levels(melted_correlation_matrix$X2)
  )
)

#Correlation plot
ggplot(melted_correlation_matrix, aes(x = X1, y = X2, fill = value)) +
  geom_tile() +
  geom_text(aes(x = X1, y = X2, label = value), size = 3) +
  guides(fill = FALSE) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, size = 6, vjust = 0.2),
        axis.text.y = element_text(size = 5, hjust = 0.2),
        axis.title = element_blank())
```



Correlation Plot

The level of Life expectancy is highly and positively related to the Income composition of resources, Schooling and BMI.

Life expectancy is related

1. Strongly Positive:

- Schooling
- Income composition of resources
- BMI

2. Weakly Positive:

- Total expenditure
- Polio
- Percentage Expenditure
- Hepatitis B

- GDP
- Diphtheria
- Alcohol .

3. Weakly Negative:

- under-five deaths
- thinness 1-19 years
- thinness 5-9 years
- Measles
- infant deaths

4. Strongly Negative:

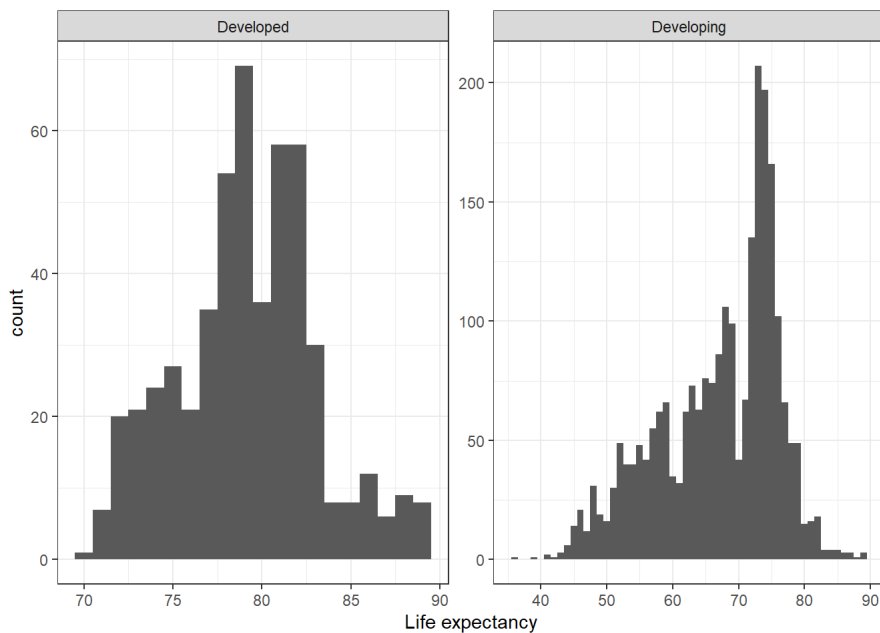
- Population
- HIV/AIDS
- Adult Mortality

Data Visualization

Histograms

A histogram of the `Life expectancy` variable, grouped by the `Status` variable provides an overview of the distribution of the data across the developed and developing nations. The developed nations histogram is more symmetrical as compared to the developing nations that is skewed.

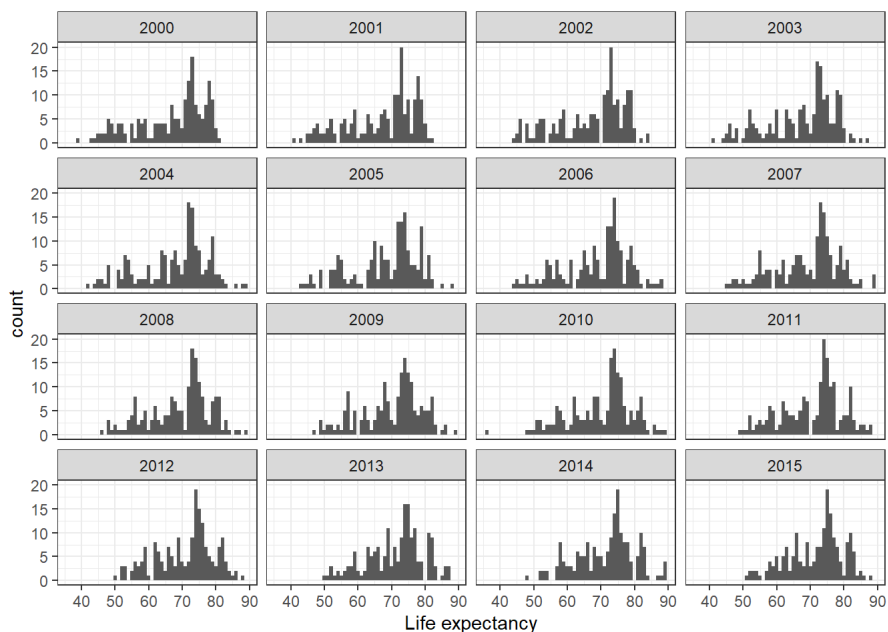
```
Life_Expectancy_Data %>%
  ggplot(aes(x = `Life expectancy`)) +
  geom_histogram(binwidth = 1) +
  facet_wrap(Status~., scales = "free") +
  theme_bw()
```



Life expectancy Distribution by Status

The distribution of the `Life expectancy` variable across the various years is plotted and is left-skewed across the years.

```
Life_Expectancy_Data %>%
  ggplot(aes(x = `Life expectancy`)) +
  geom_histogram(binwidth = 1) +
  facet_wrap(Year~.,) +
  theme_bw()
```

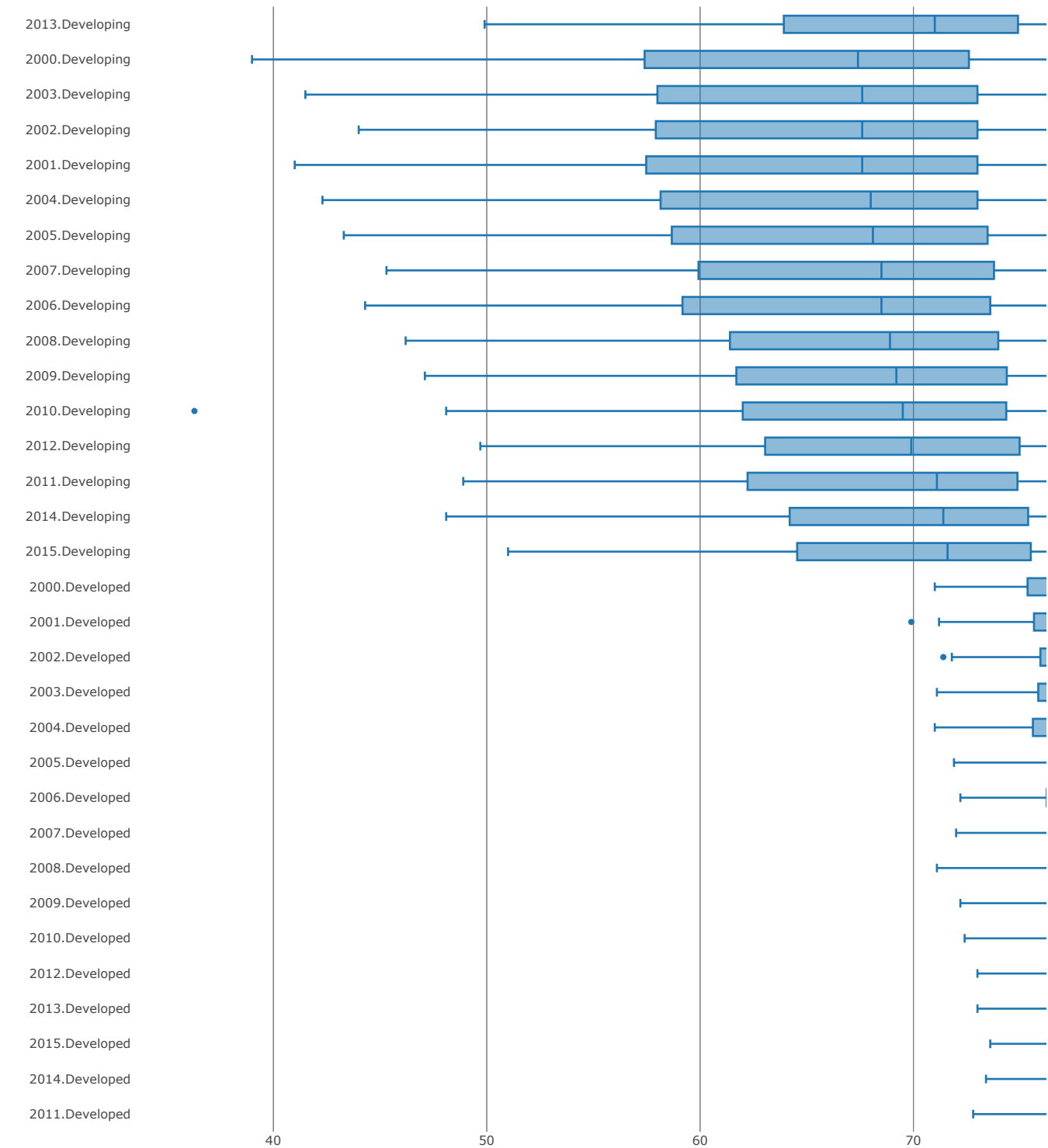


Life expectancy Distribution by Year

Box plot

A box plot display the distribution of the data, pointing out the upper, middle and lower quantile. The interaction between the Year and Status of country variable are made and box plots of Life Expectancy graphed, ordered by the median of the Life Expectancy .

```
library(plotly)
dat <- Life_Expectancy_Data %>%
  mutate(inter = interaction(Year, Status))
# interaction levels sorted by median life expectancy
levelS <- dat %>%
  group_by(inter) %>%
  summarise(m = median(`Life expectancy`)) %>%
  arrange(desc(m)) %>%
  pull(inter)
plot_ly(dat, x = `Life expectancy`, y = ~factor(inter, levelS)) %>%
  add_boxplot() %>%
  layout(yaxis = list(title = ""))
```



Year-Status Life Expectancy Box plot

Scatter plots

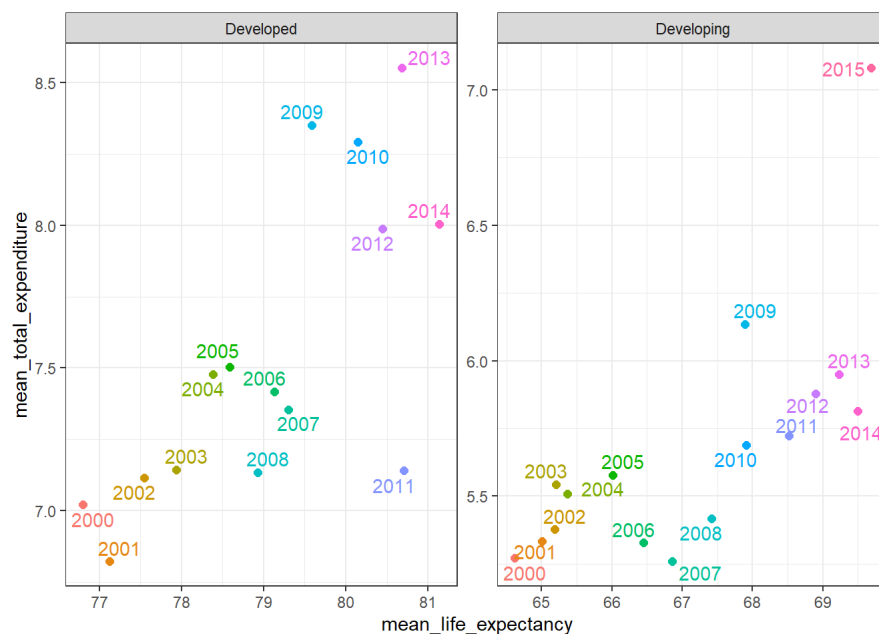
The scatter plots on Life Expenditure against the Total Expenditure, Percentage Expenditure and Schooling show the trends observed across the extracted variables. The plots are grouped according to the Status variable to bring out the distinctions among countries.

Life expectancy and total expenditure.

The graph of the mean Total expenditure vs mean Life expectancy show the discord in expenditure and how they relate. Developed countries have higher mean Life expectancy and also higher mean Total expenditure as compared to their Developing counterparts.

```
#Life Expectancy total Expenditure
```

```
Life_Expectancy_Data %>% group_by(Status, Year) %>%
  summarise(
    mean_total_expenditure = mean(`Total expenditure`, na.rm = TRUE),
    mean_life_expectancy = mean(`Life expectancy`, na.rm = TRUE)
  ) %>%
  ggplot(aes(x = mean_life_expectancy, y = mean_total_expenditure,
    color = Year)) + geom_point(size = 2) +
  ggrepel::geom_text_repel(aes(label = Year)) +
  facet_wrap(Status~., scales = "free") +
  guides(color = FALSE) + theme_bw() +
  theme(legend.key = element_blank())
```

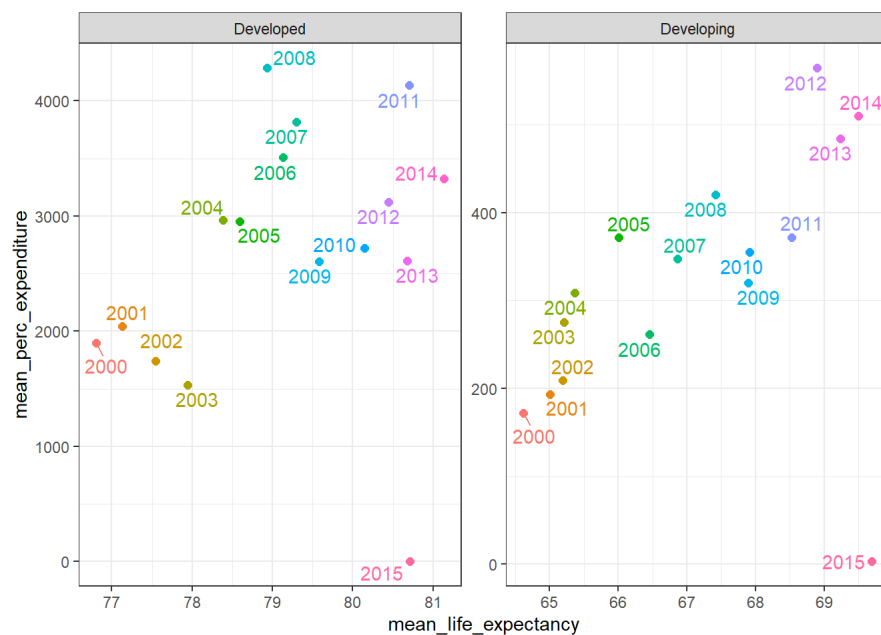


Life Expectancy vs total Expenditure

Life expectancy and percentage expenditure.

The graph of the mean percentage expenditure vs mean Life expectancy show how they relate. Developed countries have higher mean Life expectancy and also higher mean percentage expenditure as compared to their Developing counterparts.

```
#Life Expectancy % expenditure
Life_Expectancy_Data %>% group_by(Status, Year) %>%
  summarise(
    mean_perc_expenditure = mean('percentage expenditure', na.rm = TRUE),
    mean_life_expectancy = mean('Life expectancy', na.rm = TRUE)
  ) %>%
  ggplot(aes(x = mean_life_expectancy, y = mean_perc_expenditure,
             color = Year)) + geom_point(size = 2) +
  ggrepel::geom_text_repel(aes(label = Year)) +
  facet_wrap(Status~., scales = "free") +
  guides(color = FALSE) + theme_bw() +
  theme(legend.key = element_blank())
```

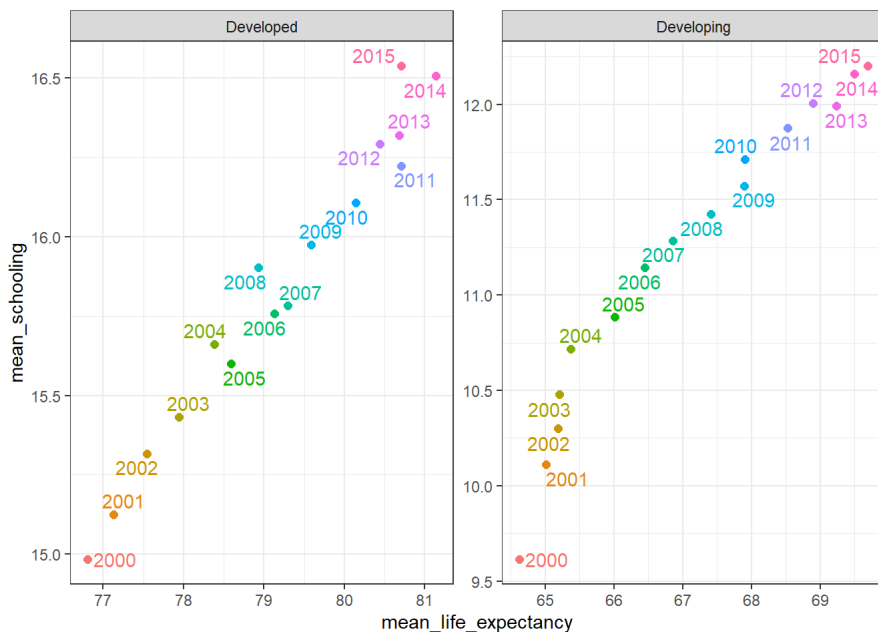


Life Expectancy vs Percentage Expenditure

Life Expectancy and Schooling.

The graph of the mean `Schooling` vs mean `Life expectancy` show how they relate. Developed countries have higher mean `Life expectancy` and also higher mean `Schooling` as compared to their Developing counterparts. The visualization can be interpreted as the better the education provided, especially on healthy living within the developed countries, the better the rates of life expectancy as compared to the developing countries.

```
#Life Expectancy Schooling
Life_Expectancy_Data %>% group_by(Status, Year) %>%
  summarise(
    mean_schooling = mean(Schooling, na.rm = TRUE),
    mean_life_expectancy = mean('Life expectancy', na.rm = TRUE)
  ) %>%
  ggplot(aes(x = mean_life_expectancy, y = mean_schooling,
    color = Year)) + geom_point(size = 2) +
  ggrepel::geom_text_repel(aes(label = Year)) +
  facet_wrap(Status~., scales = "free") +
  guides(color = FALSE) + theme_bw() +
  theme(legend.key = element_blank())
```



Methods

A multiple regression is undertaken on the data, with the `Life expectancy` as the response variable and the rest of the variables as the predictors. An analysis of variance of the model from the regression is made.

Regression

The outcome for our regression model is the `Life expectancy`. A few of the explanatory variables `Country`, `Year` and `Status` are categorical, and with 193, 16 and 2 levels respectively. The rest of the variables are numeric variables.

```
library(broom)
model <- lm(`Life expectancy`~, data = Life_Expectancy_Data)
t(glance(model))
```

```
##               [,1]
## r.squared      0.9678564
## adj.r.squared  0.9642801
## sigma         1.6625758
## statistic      270.6288482
## p.value        0.0000000
## df            165.0000000
## logLik        -3090.6475570
## AIC           6515.2951140
## BIC           7418.4184759
## deviance      4099.2468645
## df.residual    1483.0000000
## nobs          1649.0000000
```

The fraction of variation of the dependent variable explained by the regression line, the R squared (R^2) is at 0.96 for the model.

The regression table from the model is:

```
knitr::kable( moderndive::get_regression_table(model = model))
```

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	53.091	0.979	54.247	0.000	51.172	55.011
CountryAlbania	16.638	0.860	19.355	0.000	14.952	18.324
CountryAlgeria	14.564	0.855	17.031	0.000	12.887	16.241
CountryAngola	-6.267	0.825	-7.600	0.000	-7.885	-4.649
CountryArgentina	15.777	1.076	14.661	0.000	13.666	17.888
CountryArmenia	15.105	0.841	17.957	0.000	13.455	16.756
CountryAustralia	21.721	1.315	16.512	0.000	19.140	24.301
CountryAustria	22.705	1.081	20.998	0.000	20.584	24.826
CountryAzerbaijan	12.656	0.825	15.334	0.000	11.037	14.275
CountryBangladesh	11.195	0.707	15.836	0.000	9.808	12.582
CountryBelarus	11.476	1.017	11.289	0.000	9.482	13.471
CountryBelgium	21.539	1.127	19.104	0.000	19.327	23.750
CountryBelize	10.937	0.852	12.840	0.000	9.266	12.608
CountryBenin	0.506	0.696	0.727	0.467	-0.859	1.872
CountryBhutan	7.305	0.678	10.783	0.000	5.976	8.634
CountryBosnia and Herzegovina	17.379	0.896	19.394	0.000	15.622	19.137
CountryBotswana	2.242	0.817	2.745	0.006	0.640	3.844
CountryBrazil	14.121	0.929	15.197	0.000	12.298	15.943
CountryBulgaria	14.493	0.956	15.155	0.000	12.617	16.369
CountryBurkina Faso	0.780	0.805	0.969	0.333	-0.798	2.358
CountryBurundi	-0.782	0.735	-1.064	0.288	-2.225	0.660
CountryCabo Verde	13.835	0.784	17.644	0.000	12.297	15.373
CountryCambodia	7.268	0.757	9.598	0.000	5.783	8.754
CountryCameroon	-1.067	0.777	-1.374	0.170	-2.591	0.457
CountryCanada	23.070	1.108	20.830	0.000	20.898	25.243
CountryCentral African Republic	-4.770	0.880	-5.422	0.000	-6.496	-3.045
CountryChad	-3.879	0.846	-4.588	0.000	-5.538	-2.221
CountryChile	20.355	1.051	19.360	0.000	18.292	22.417
CountryChina	14.733	1.209	12.188	0.000	12.362	17.104
CountryColombia	14.956	0.844	17.715	0.000	13.300	16.612
CountryComoros	3.645	0.733	4.976	0.000	2.208	5.082
CountryCosta Rica	20.150	0.877	22.979	0.000	18.430	21.870
CountryCroatia	17.551	1.074	16.341	0.000	15.444	19.658
CountryCyprus	20.956	0.988	21.209	0.000	19.018	22.895
CountryDjibouti	6.166	0.855	7.212	0.000	4.489	7.843
CountryDominican Republic	14.334	0.855	16.764	0.000	12.657	16.011
CountryEcuador	16.203	0.878	18.447	0.000	14.480	17.926
CountryEl Salvador	13.303	0.854	15.570	0.000	11.627	14.979
CountryEquatorial Guinea	-0.140	1.763	-0.079	0.937	-3.599	3.319
CountryEritrea	5.038	0.792	6.357	0.000	3.483	6.592
CountryEstonia	15.281	1.079	14.164	0.000	13.164	17.397
CountryEthiopia	6.350	0.859	7.393	0.000	4.666	8.035
CountryFiji	9.420	0.894	10.532	0.000	7.666	11.174

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
CountryFrance	23.546	1.115	21.126	0.000	21.360	25.732
CountryGabon	6.254	0.865	7.231	0.000	4.557	7.950
CountryGeorgia	15.190	0.855	17.758	0.000	13.512	16.868
CountryGermany	22.141	1.138	19.459	0.000	19.909	24.373
CountryGhana	4.033	0.710	5.677	0.000	2.640	5.427
CountryGreece	22.213	1.063	20.893	0.000	20.127	24.298
CountryGuatemala	14.928	0.850	17.570	0.000	13.261	16.594
CountryGuinea	0.098	0.784	0.125	0.901	-1.440	1.635
CountryGuinea-Bissau	0.307	0.868	0.354	0.723	-1.395	2.010
CountryGuyana	8.057	0.787	10.236	0.000	6.513	9.601
CountryHaiti	4.479	1.319	3.397	0.001	1.893	7.066
CountryHonduras	15.237	0.801	19.029	0.000	13.667	16.808
CountryIndia	6.049	2.858	2.116	0.034	0.442	11.655
CountryIndonesia	8.557	0.831	10.298	0.000	6.927	10.187
CountryIraq	11.716	0.780	15.012	0.000	10.185	13.247
CountryIreland	22.624	1.334	16.954	0.000	20.006	25.241
CountryIsrael	21.613	1.031	20.959	0.000	19.591	23.636
CountryItaly	23.064	1.073	21.491	0.000	20.958	25.169
CountryJamaica	16.137	0.883	18.285	0.000	14.405	17.868
CountryJordan	13.924	0.861	16.170	0.000	12.235	15.614
CountryKazakhstan	7.687	0.924	8.323	0.000	5.875	9.498
CountryKenya	2.393	0.716	3.341	0.001	0.988	3.798
CountryKiribati	7.227	0.859	8.412	0.000	5.542	8.913
CountryLatvia	14.688	1.010	14.541	0.000	12.707	16.670
CountryLebanon	15.300	0.851	17.975	0.000	13.630	16.969
CountryLesotho	-2.929	0.825	-3.549	0.000	-4.548	-1.310
CountryLiberia	2.788	0.826	3.376	0.001	1.168	4.408
CountryLithuania	13.813	1.065	12.973	0.000	11.724	15.901
CountryLuxembourg	22.398	1.068	20.970	0.000	20.303	24.493
CountryMadagascar	5.478	0.700	7.830	0.000	4.105	6.850
CountryMalawi	-2.568	0.773	-3.320	0.001	-4.085	-1.050
CountryMalaysia	14.390	0.792	18.167	0.000	12.837	15.944
CountryMaldives	16.027	0.722	22.192	0.000	14.611	17.444
CountryMali	-0.658	0.743	-0.885	0.376	-2.115	0.800
CountryMalta	22.011	1.004	21.933	0.000	20.042	23.979
CountryMauritania	5.460	0.748	7.296	0.000	3.992	6.928
CountryMauritius	13.378	0.826	16.188	0.000	11.757	14.999
CountryMexico	17.183	0.867	19.829	0.000	15.483	18.883
CountryMongolia	7.228	0.848	8.521	0.000	5.564	8.892
CountryMontenegro	15.473	0.992	15.602	0.000	13.527	17.418
CountryMorocco	13.697	0.734	18.673	0.000	12.258	15.136
CountryMozambique	0.121	0.762	0.159	0.874	-1.374	1.617
CountryMyanmar	6.125	0.679	9.019	0.000	4.793	7.457
CountryNamibia	6.642	0.945	7.032	0.000	4.789	8.495
CountryNepal	7.794	0.687	11.345	0.000	6.446	9.142
CountryNetherlands	19.936	1.424	13.999	0.000	17.142	22.729

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
CountryNicaragua	15.540	0.812	19.126	0.000	13.946	17.133
CountryNiger	5.284	0.979	5.397	0.000	3.363	7.204
CountryNigeria	1.140	1.677	0.680	0.497	-2.149	4.429
CountryPakistan	5.548	1.093	5.074	0.000	3.403	7.693
CountryPanama	18.112	0.891	20.319	0.000	16.363	19.860
CountryPapua New Guinea	4.913	0.771	6.370	0.000	3.400	6.426
CountryParaguay	14.434	0.869	16.616	0.000	12.730	16.138
CountryPeru	15.122	0.913	16.558	0.000	13.330	16.913
CountryPhilippines	9.000	0.825	10.903	0.000	7.381	10.619
CountryPoland	16.645	1.013	16.438	0.000	14.659	18.632
CountryPortugal	21.202	1.085	19.534	0.000	19.073	23.331
CountryRomania	15.416	0.933	16.522	0.000	13.586	17.246
CountryRussian Federation	9.055	0.963	9.401	0.000	7.165	10.944
CountryRwanda	4.128	0.743	5.553	0.000	2.670	5.587
CountrySamoa	15.402	0.890	17.310	0.000	13.656	17.147
CountrySao Tome and Principe	7.596	0.762	9.968	0.000	6.102	9.091
CountrySenegal	6.338	0.753	8.421	0.000	4.862	7.814
CountrySerbia	15.412	0.961	16.032	0.000	13.526	17.298
CountrySeychelles	13.590	0.836	16.257	0.000	11.950	15.230
CountrySierra Leone	-9.318	0.782	-11.918	0.000	-10.852	-7.784
CountrySolomon Islands	10.448	0.786	13.285	0.000	8.905	11.991
CountrySouth Africa	4.625	0.830	5.572	0.000	2.997	6.253
CountrySpain	23.015	1.096	20.994	0.000	20.865	25.166
CountrySri Lanka	13.325	0.802	16.625	0.000	11.753	14.898
CountrySuriname	12.386	0.880	14.075	0.000	10.660	14.112
CountrySwaziland	3.489	0.909	3.837	0.000	1.705	5.272
CountrySweden	21.724	1.279	16.987	0.000	19.216	24.233
CountrySyrian Arab Republic	15.684	0.861	18.225	0.000	13.996	17.372
CountryTajikistan	8.531	0.782	10.905	0.000	6.996	10.065
CountryThailand	14.219	0.797	17.852	0.000	12.657	15.781
CountryTimor-Leste	6.948	0.869	7.995	0.000	5.243	8.652
CountryTogo	-0.251	0.848	-0.296	0.767	-1.913	1.412
CountryTonga	13.648	0.948	14.391	0.000	11.788	15.508
CountryTrinidad and Tobago	12.844	0.833	15.422	0.000	11.210	14.477
CountryTunisia	14.883	0.864	17.230	0.000	13.189	16.578
CountryTurkey	14.888	0.816	18.254	0.000	13.288	16.488
CountryTurkmenistan	6.793	0.808	8.407	0.000	5.208	8.378
CountryUganda	1.770	0.790	2.242	0.025	0.221	3.319
CountryUkraine	11.424	0.942	12.132	0.000	9.577	13.271
CountryUruguay	17.047	0.987	17.270	0.000	15.111	18.984
CountryUzbekistan	9.426	0.817	11.539	0.000	7.824	11.028
CountryVanuatu	13.741	0.801	17.146	0.000	12.169	15.313
CountryZambia	1.456	0.807	1.805	0.071	-0.126	3.038
CountryZimbabwe	-0.545	0.801	-0.680	0.497	-2.117	1.027
Year2001	0.237	0.297	0.799	0.425	-0.345	0.820
Year2002	0.098	0.286	0.342	0.732	-0.463	0.659

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
Year2003	0.161	0.280	0.574	0.566	-0.389	0.711
Year2004	0.254	0.279	0.910	0.363	-0.293	0.800
Year2005	0.727	0.280	2.599	0.009	0.178	1.275
Year2006	1.046	0.284	3.688	0.000	0.490	1.602
Year2007	1.250	0.285	4.386	0.000	0.691	1.809
Year2008	1.513	0.291	5.199	0.000	0.942	2.084
Year2009	1.782	0.295	6.050	0.000	1.205	2.360
Year2010	1.937	0.299	6.478	0.000	1.350	2.523
Year2011	2.330	0.309	7.553	0.000	1.725	2.936
Year2012	2.450	0.315	7.787	0.000	1.833	3.067
Year2013	2.574	0.319	8.066	0.000	1.948	3.200
Year2014	2.659	0.327	8.133	0.000	2.018	3.301
Year2015	5.536	1.262	4.388	0.000	3.061	8.011
StatusDeveloping	NA	NA	NA	NA	NA	NA
Adult Mortality	-0.001	0.001	-1.178	0.239	-0.002	0.000
infant deaths	0.048	0.016	3.099	0.002	0.018	0.079
Alcohol	-0.073	0.032	-2.268	0.024	-0.135	-0.010
percentage expenditure	0.000	0.000	-0.708	0.479	0.000	0.000
Hepatitis B	0.003	0.002	1.314	0.189	-0.002	0.008
Measles	0.000	0.000	-0.884	0.377	0.000	0.000
BMI	-0.002	0.003	-0.452	0.652	-0.008	0.005
under-five deaths	-0.036	0.011	-3.291	0.001	-0.058	-0.015
Polio	-0.001	0.003	-0.199	0.843	-0.006	0.005
Total expenditure	-0.020	0.027	-0.740	0.460	-0.072	0.032
Diphtheria	0.000	0.003	0.157	0.875	-0.005	0.006
HIV/AIDS	-0.301	0.016	-19.008	0.000	-0.332	-0.270
GDP	0.000	0.000	0.850	0.395	0.000	0.000
Population	0.000	0.000	-0.174	0.862	0.000	0.000
thinness 1-19 years	0.008	0.033	0.258	0.796	-0.056	0.073
thinness 5-9 years	0.069	0.031	2.192	0.029	0.007	0.130
Income composition of resources	0.885	0.597	1.482	0.139	-0.286	2.056
Schooling	0.281	0.078	3.594	0.000	0.128	0.434

Based on the estimate column of the regression table, Afghanistan Country was the "baseline for comparison" group, therefore, the intercept term corresponds to the life expectancy for the Afghanistan country. The other values of estimate correspond to "offsets" relative to the baseline group.

Analysis of Variance

Stepwise selection

We use the `step()` function to explore a variety of variables for our model with only the significant variables.

```
model <- lm(`Life expectancy`~., data = na.omit(Life_Expectancy_Data))  
model2 <- step(model, direction = "both")
```

```

## Start: AIC=1833.64
## `Life expectancy` ~ Country + Year + Status + `Adult Mortality` +
##   `infant deaths` + Alcohol + `percentage expenditure` + `Hepatitis B` +
##   Measles + BMI + `under-five deaths` + Polio + `Total expenditure` +
##   Diphtheria + `HIV/AIDS` + GDP + Population + `thinness 1-19 years` +
##   `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##
## Step: AIC=1833.64
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
##   BMI + `under-five deaths` + Polio + `Total expenditure` +
##   Diphtheria + `HIV/AIDS` + GDP + Population + `thinness 1-19 years` +
##   `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##
##              Df Sum of Sq    RSS    AIC
## - Diphtheria      1      0.1 4099.3 1831.7
## - Population      1      0.1 4099.3 1831.7
## - Polio            1      0.1 4099.4 1831.7
## - `thinness 1-19 years` 1      0.2 4099.4 1831.7
## - BMI             1      0.6 4099.8 1831.9
## - `percentage expenditure` 1     1.4 4100.6 1832.2
## - `Total expenditure` 1     1.5 4100.8 1832.2
## - GDP             1     2.0 4101.2 1832.4
## - Measles         1     2.2 4101.4 1832.5
## - `Adult Mortality` 1     3.8 4103.1 1833.2
## - `Hepatitis B`    1     4.8 4104.0 1833.6
## <none>                                4099.2 1833.6
## - `Income composition of resources` 1     6.1 4105.3 1834.1
## - `thinness 5-9 years` 1    13.3 4112.5 1837.0
## - Alcohol          1    14.2 4113.5 1837.3
## - `infant deaths`  1    26.5 4125.8 1842.3
## - `under-five deaths` 1    29.9 4129.2 1843.6
## - Schooling        1    35.7 4134.9 1845.9
## - Year             15   504.9 4604.1 1995.2
## - `HIV/AIDS`       1    998.7 5097.9 2191.2
## - Country          132 16516.5 20615.7 4233.2
##
## Step: AIC=1831.66
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
##   BMI + `under-five deaths` + Polio + `Total expenditure` +
##   `HIV/AIDS` + GDP + Population + `thinness 1-19 years` +
##   `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##
##              Df Sum of Sq    RSS    AIC
## - Polio            1      0.1 4099.4 1829.7
## - Population      1      0.1 4099.4 1829.7
## - `thinness 1-19 years` 1      0.2 4099.5 1829.7
## - BMI             1      0.6 4099.9 1829.9
## - `percentage expenditure` 1     1.4 4100.7 1830.2
## - `Total expenditure` 1     1.5 4100.8 1830.3
## - GDP             1     2.0 4101.3 1830.5
## - Measles         1     2.2 4101.5 1830.5
## - `Adult Mortality` 1     3.8 4103.1 1831.2
## <none>                                4099.3 1831.7
## - `Income composition of resources` 1     6.2 4105.6 1832.2
## - `Hepatitis B`    1     6.3 4105.6 1832.2
## + Diphtheria      1      0.1 4099.2 1833.6
## - `thinness 5-9 years` 1    13.3 4112.6 1835.0
## - Alcohol          1    14.2 4113.5 1835.4
## - `infant deaths`  1    26.8 4126.1 1840.4
## - `under-five deaths` 1    30.2 4129.6 1841.8
## - Schooling        1    35.8 4135.1 1844.0
## - Year             15   505.1 4604.4 1993.3
## - `HIV/AIDS`       1   1001.4 5100.7 2190.1
## - Country          132 16585.1 20684.5 4236.7
##
## Step: AIC=1829.69
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
##   BMI + `under-five deaths` + `Total expenditure` + `HIV/AIDS` +
##   GDP + Population + `thinness 1-19 years` + `thinness 5-9 years` +
##   `Income composition of resources` + Schooling
##
##
##              Df Sum of Sq    RSS    AIC
## - Population      1      0.1 4099.5 1827.7
## - `thinness 1-19 years` 1      0.2 4099.5 1827.8

```

```

## - BMI 1 0.6 4099.9 1827.9
## - `percentage expenditure` 1 1.4 4100.8 1828.2
## - `Total expenditure` 1 1.5 4100.9 1828.3
## - GDP 1 2.0 4101.4 1828.5
## - Measles 1 2.2 4101.5 1828.6
## - `Adult Mortality` 1 3.8 4103.2 1829.2
## <none> 4099.4 1829.7
## - `Income composition of resources` 1 6.2 4105.6 1830.2
## - `Hepatitis B` 1 6.5 4105.8 1830.3
## + Polio 1 0.1 4099.3 1831.7
## + Diphtheria 1 0.0 4099.4 1831.7
## - `thinness 5-9 years` 1 13.4 4112.7 1833.1
## - Alcohol 1 14.2 4113.5 1833.4
## - `infant deaths` 1 26.8 4126.2 1838.5
## - `under-five deaths` 1 30.3 4129.6 1839.8
## - Schooling 1 35.7 4135.1 1842.0
## - Year 15 505.5 4604.9 1991.4
## - `HIV/AIDS` 1 1001.5 5100.8 2188.1
## - Country 132 16649.4 20748.8 4239.8
##
## Step: AIC=1827.72
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
## Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
## BMI + `under-five deaths` + `Total expenditure` + `HIV/AIDS` +
## GDP + `thinness 1-19 years` + `thinness 5-9 years` + `Income composition of resources` +
## Schooling
##
## Df Sum of Sq RSS AIC
## - `thinness 1-19 years` 1 0.1 4099.6 1825.8
## - BMI 1 0.6 4100.0 1826.0
## - `percentage expenditure` 1 1.4 4100.8 1826.3
## - `Total expenditure` 1 1.5 4101.0 1826.3
## - GDP 1 2.0 4101.4 1826.5
## - Measles 1 2.2 4101.7 1826.6
## - `Adult Mortality` 1 3.8 4103.3 1827.3
## <none> 4099.5 1827.7
## - `Income composition of resources` 1 6.2 4105.7 1828.2
## - `Hepatitis B` 1 6.4 4105.9 1828.3
## + Population 1 0.1 4099.4 1829.7
## + Polio 1 0.1 4099.4 1829.7
## + Diphtheria 1 0.0 4099.4 1829.7
## - `thinness 5-9 years` 1 13.6 4113.0 1831.2
## - Alcohol 1 14.2 4113.7 1831.4
## - `infant deaths` 1 28.1 4127.6 1837.0
## - `under-five deaths` 1 31.1 4130.6 1838.2
## - Schooling 1 35.6 4135.1 1840.0
## - Year 15 505.4 4604.9 1989.4
## - `HIV/AIDS` 1 1002.2 5101.7 2186.4
## - Country 132 16650.2 20749.7 4237.9
##
## Step: AIC=1825.78
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
## Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
## BMI + `under-five deaths` + `Total expenditure` + `HIV/AIDS` +
## GDP + `thinness 5-9 years` + `Income composition of resources` +
## Schooling
##
## Df Sum of Sq RSS AIC
## - BMI 1 0.6 4100.2 1824.0
## - `percentage expenditure` 1 1.4 4101.0 1824.3
## - `Total expenditure` 1 1.5 4101.1 1824.4
## - GDP 1 2.0 4101.6 1824.6
## - Measles 1 2.2 4101.8 1824.7
## - `Adult Mortality` 1 3.9 4103.5 1825.3
## <none> 4099.6 1825.8
## - `Income composition of resources` 1 6.2 4105.8 1826.3
## - `Hepatitis B` 1 6.5 4106.1 1826.4
## + `thinness 1-19 years` 1 0.1 4099.5 1827.7
## + Population 1 0.1 4099.5 1827.8
## + Polio 1 0.1 4099.6 1827.8
## + Diphtheria 1 0.0 4099.6 1827.8
## - Alcohol 1 14.2 4113.8 1829.5
## - `thinness 5-9 years` 1 22.4 4122.0 1832.8
## - `infant deaths` 1 28.0 4127.6 1835.0
## - `under-five deaths` 1 31.0 4130.6 1836.2
## - Schooling 1 35.6 4135.2 1838.0
## - Year 15 505.3 4605.0 1987.5
## - `HIV/AIDS` 1 1002.4 5102.1 2184.5
## - Country 132 16650.1 20749.7 4235.9
##
## Step: AIC=1824.01

```

```

## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
##   `under-five deaths` + `Total expenditure` + `HIV/AIDS` +
##   GDP + `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##
##               Df Sum of Sq    RSS    AIC
## - `Total expenditure`      1      1.4  4101.6 1822.6
## - `percentage expenditure`  1      1.5  4101.7 1822.6
## - GDP                      1      2.1  4102.3 1822.9
## - Measles                  1      2.3  4102.4 1822.9
## - `Adult Mortality`       1      3.9  4104.0 1823.6
## <none>                     4100.2 1824.0
## - `Income composition of resources` 1      6.0  4106.2 1824.4
## - `Hepatitis B`          1      6.3  4106.4 1824.5
## + BMI                    1      0.6  4099.6 1825.8
## + `thinness 1-19 years`  1      0.1  4100.0 1826.0
## + Population             1      0.1  4100.1 1826.0
## + Polio                  1      0.0  4100.1 1826.0
## + Diphtheria             1      0.0  4100.1 1826.0
## - Alcohol                1     14.3  4114.5 1827.8
## - `thinness 5-9 years`    1     22.6  4122.8 1831.1
## - `infant deaths`        1     28.0  4128.2 1833.2
## - `under-five deaths`    1     31.0  4131.2 1834.4
## - Schooling              1     35.6  4135.8 1836.3
## - Year                   15    505.6  4605.8 1985.8
## - `HIV/AIDS`             1    1007.4  5107.6 2184.3
## - Country                132   16967.8 21067.9 4259.0
##
## Step: AIC=1822.58
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `percentage expenditure` + `Hepatitis B` + Measles +
##   `under-five deaths` + `HIV/AIDS` + GDP + `thinness 5-9 years` +
##   `Income composition of resources` + Schooling
##
##
##               Df Sum of Sq    RSS    AIC
## - `percentage expenditure`  1      1.6  4103.2 1821.2
## - Measles                  1      2.1  4103.7 1821.4
## - GDP                      1      2.2  4103.8 1821.5
## - `Adult Mortality`       1      3.9  4105.5 1822.1
## <none>                     4101.6 1822.6
## - `Income composition of resources` 1      6.0  4107.6 1823.0
## - `Hepatitis B`          1      6.1  4107.7 1823.0
## + `Total expenditure`     1      1.4  4100.2 1824.0
## + BMI                    1      0.5  4101.1 1824.4
## + `thinness 1-19 years`  1      0.1  4101.5 1824.5
## + Population             1      0.1  4101.5 1824.6
## + Polio                  1      0.1  4101.5 1824.6
## + Diphtheria             1      0.0  4101.6 1824.6
## - Alcohol                1     14.2  4115.8 1826.3
## - `thinness 5-9 years`    1     22.5  4124.1 1829.6
## - `infant deaths`        1     28.1  4129.7 1831.9
## - `under-five deaths`    1     31.2  4132.8 1833.1
## - Schooling              1     35.8  4137.4 1834.9
## - Year                   15    504.2  4605.8 1983.8
## - `HIV/AIDS`             1    1006.0  5107.6 2182.3
## - Country                132   17052.4 21154.0 4263.7
##
## Step: AIC=1821.21
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `Hepatitis B` + Measles + `under-five deaths` +
##   `HIV/AIDS` + GDP + `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##
##               Df Sum of Sq    RSS    AIC
## - GDP                      1      0.8  4104.0 1819.5
## - Measles                  1      2.1  4105.3 1820.1
## - `Adult Mortality`       1      3.8  4107.0 1820.8
## <none>                     4103.2 1821.2
## - `Income composition of resources` 1      6.1  4109.2 1821.6
## - `Hepatitis B`          1      6.2  4109.4 1821.7
## + `percentage expenditure`  1      1.6  4101.6 1822.6
## + `Total expenditure`     1      1.5  4101.7 1822.6
## + BMI                    1      0.6  4102.6 1823.0
## + `thinness 1-19 years`  1      0.2  4103.0 1823.2
## + Population             1      0.1  4103.1 1823.2
## + Polio                  1      0.1  4103.1 1823.2
## + Diphtheria             1      0.0  4103.1 1823.2
## - Alcohol                1     13.7  4116.8 1824.7
## - `thinness 5-9 years`    1     22.4  4125.6 1828.2
## - `infant deaths`        1     27.7  4130.9 1830.3

```



```

## - `under-five deaths`      1      30.8  4133.9 1831.5
## - Schooling                 1      36.5  4139.7 1833.8
## - Year                     15     506.9  4610.1 1983.3
## - `HIV/AIDS`              1    1004.5  5107.7 2180.3
## - Country                 132   17097.2 21200.4 4265.3
##
## Step: AIC=1819.54
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `Hepatitis B` + Measles + `under-five deaths` +
##   `HIV/AIDS` + `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##                                     Df Sum of Sq    RSS    AIC
## - Measles                        1      2.1  4106.1 1818.4
## - `Adult Mortality`              1      3.8  4107.8 1819.1
## <none>                           1      4104.0 1819.5
## - `Income composition of resources` 1      5.9  4109.9 1819.9
## - `Hepatitis B`                 1      6.1  4110.1 1820.0
## + `Total expenditure`           1      1.4  4102.6 1821.0
## + GDP                           1      0.8  4103.2 1821.2
## + BMI                           1      0.6  4103.4 1821.3
## + `percentage expenditure`       1      0.2  4103.8 1821.5
## + `thinness 1-19 years`         1      0.2  4103.8 1821.5
## + Population                    1      0.1  4103.9 1821.5
## + Polio                         1      0.1  4103.9 1821.5
## + Diphtheria                    1      0.0  4104.0 1821.5
## - Alcohol                       1     13.5  4117.5 1823.0
## - `thinness 5-9 years`          1     22.4  4126.4 1826.5
## - `infant deaths`               1     27.7  4131.7 1828.6
## - `under-five deaths`           1     30.7  4134.7 1829.8
## - Schooling                     1     36.0  4140.0 1831.9
## - Year                          15    530.5  4634.5 1990.0
## - `HIV/AIDS`                   1    1003.7  5107.7 2178.3
## - Country                       132   17992.8 22096.8 4331.6
##
## Step: AIC=1818.4
## `Life expectancy` ~ Country + Year + `Adult Mortality` + `infant deaths` +
##   Alcohol + `Hepatitis B` + `under-five deaths` + `HIV/AIDS` +
##   `thinness 5-9 years` + `Income composition of resources` +
##   Schooling
##
##                                     Df Sum of Sq    RSS    AIC
## - `Adult Mortality`              1      3.8  4109.9 1817.9
## <none>                           1      4106.1 1818.4
## - `Income composition of resources` 1      5.8  4111.9 1818.7
## - `Hepatitis B`                 1      6.2  4112.3 1818.9
## + Measles                       1      2.1  4104.0 1819.5
## + `Total expenditure`           1      1.3  4104.8 1819.9
## + GDP                           1      0.8  4105.3 1820.1
## + BMI                           1      0.6  4105.5 1820.1
## + `percentage expenditure`       1      0.2  4105.9 1820.3
## + `thinness 1-19 years`         1      0.1  4106.0 1820.3
## + Population                    1      0.1  4106.0 1820.3
## + Polio                         1      0.1  4106.0 1820.4
## + Diphtheria                    1      0.0  4106.1 1820.4
## - Alcohol                       1     13.6  4119.7 1821.9
## - `thinness 5-9 years`          1     22.9  4129.0 1825.6
## - `infant deaths`               1     26.4  4132.5 1826.9
## - `under-five deaths`           1     29.5  4135.7 1828.2
## - Schooling                     1     37.0  4143.1 1831.2
## - Year                          15    532.4  4638.5 1989.4
## - `HIV/AIDS`                   1    1002.9  5109.0 2176.7
## - Country                       132   18030.7 22136.8 4332.6
##
## Step: AIC=1817.91
## `Life expectancy` ~ Country + Year + `infant deaths` + Alcohol +
##   `Hepatitis B` + `under-five deaths` + `HIV/AIDS` + `thinness 5-9 years` +
##   `Income composition of resources` + Schooling
##
##                                     Df Sum of Sq    RSS    AIC
## <none>                           1      4109.9 1817.9
## - `Hepatitis B`                 1      5.9  4115.8 1818.3
## - `Income composition of resources` 1      6.0  4115.9 1818.3
## + `Adult Mortality`              1      3.8  4106.1 1818.4
## + Measles                       1      2.1  4107.8 1819.1
## + `Total expenditure`           1      1.3  4108.6 1819.4
## + GDP                           1      0.8  4109.1 1819.6
## + BMI                           1      0.6  4109.3 1819.7
## + `percentage expenditure`       1      0.2  4109.7 1819.8
## + Population                    1      0.2  4109.7 1819.8
## + `thinness 1-19 years`         1      0.2  4109.7 1819.8

```

## + Polio	1	0.1	4109.8	1819.9
## + Diphtheria	1	0.0	4109.9	1819.9
## - Alcohol	1	13.7	4123.6	1821.4
## - `thinness 5-9 years`	1	23.1	4133.0	1825.1
## - `infant deaths`	1	26.3	4136.2	1826.4
## - `under-five deaths`	1	29.4	4139.3	1827.7
## - Schooling	1	37.3	4147.2	1830.8
## - Year	15	538.9	4648.8	1991.1
## - `HIV/AIDS`	1	1051.5	5161.4	2191.6
## - Country	132	22622.4	26732.2	4641.6

Summary Statistics and ANOVA

```
t(glance(model2))

##           [,1]
## r.squared    0.9677730
## adj.r.squared 0.9644272
## sigma        1.6591481
## statistic     289.2554091
## p.value       0.0000000
## df           155.0000000
## logLik       -3092.7853519
## AIC           6499.5707038
## BIC           7348.6148224
## deviance     4109.8893383
## df.residual  1493.0000000
## nobs         1649.0000000
```

ANOVA

```
anova(model, model2)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1483	4099.247	NA	NA	NA	NA
2	1493	4109.889	-10	-10.64247	0.3850168	0.9536285
2 rows						

The models herein are no different as we tried to compare our original model with all variables against our model with some of the significant variables.

Results

The level of Life expectancy is highly and positively related to the Income composition of resources, Schooling and BMI. Overall, better education and higher expenditure on health creates better health care awareness by the population and better systems in place to care for the population. We were able to spot out a distinct outlier within the 2010.Developing interaction in the Life Expectancy box plots.

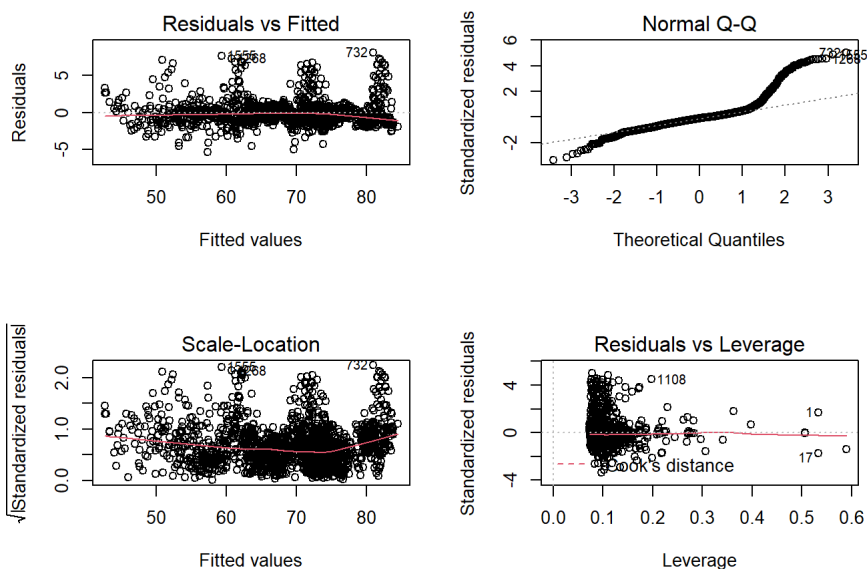
The model is 96.8% better based on the R-Squared value. The p-value of the model at 5% level shows that the model is significant. The majority of the variables' coefficients in the model are significant, with a few insignificant based on their p-values within the model. Our model is meaningful given the very low p-value on the model summary.

```
md <- moderndive::get_regression_table(model = model) %>%
  filter(p_value > 0.05) %>%
  select(term, estimate, std_error, statistic, p_value)
knitr::kable(md)
```

term	estimate	std_error	statistic	p_value
CountryBenin	0.506	0.696	0.727	0.467
CountryBurkina Faso	0.780	0.805	0.969	0.333
CountryBurundi	-0.782	0.735	-1.064	0.288
CountryCameroon	-1.067	0.777	-1.374	0.170
CountryEquatorial Guinea	-0.140	1.763	-0.079	0.937
CountryGuinea	0.098	0.784	0.125	0.901
CountryGuinea-Bissau	0.307	0.868	0.354	0.723
CountryMali	-0.658	0.743	-0.885	0.376
CountryMozambique	0.121	0.762	0.159	0.874
CountryNigeria	1.140	1.677	0.680	0.497
CountryTogo	-0.251	0.848	-0.296	0.767
CountryZambia	1.456	0.807	1.805	0.071
CountryZimbabwe	-0.545	0.801	-0.680	0.497
Year2001	0.237	0.297	0.799	0.425
Year2002	0.098	0.286	0.342	0.732
Year2003	0.161	0.280	0.574	0.566
Year2004	0.254	0.279	0.910	0.363
Adult Mortality	-0.001	0.001	-1.178	0.239
percentage expenditure	0.000	0.000	-0.708	0.479
Hepatitis B	0.003	0.002	1.314	0.189
Measles	0.000	0.000	-0.884	0.377
BMI	-0.002	0.003	-0.452	0.652
Polio	-0.001	0.003	-0.199	0.843
Total expenditure	-0.020	0.027	-0.740	0.460
Diphtheria	0.000	0.003	0.157	0.875
GDP	0.000	0.000	0.850	0.395
Population	0.000	0.000	-0.174	0.862
thinness 1-19 years	0.008	0.033	0.258	0.796
Income composition of resources	0.885	0.597	1.482	0.139

Model Residuals

```
par(mfrow = c(2,2))
plot(model)#, which = 1)
```



Residual Plots

```
par(mfrow = c(1,1))
```

The diagnostic plots show that the model is not a good one: + The points on the `Residual vs Fitted` plot are concentrated around the zero Residual point. + The residuals on the `Normal Q-Q` plot are pulling away from the line, indicating the residuals do not follow a normal distribution + The `Scale-Location` points are scattered all over + The `Residual vs Leverage` points are clustered together away from the center + Point 1294, 2717 and 2159 are sticking out on almost all plots and they could be outliers as shown below

The possible outliers that were identified from within our diagnostic plots are:

```
car::outlierTest(model)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 732  5.052871      0.0000004893  0.00080637
## 1555 4.876420      0.0000011965  0.00197190
## 1268 4.575628      0.0000051441  0.00847750
## 1218 4.564515      0.0000054203  0.00893260
## 1108 4.523084      0.0000065803  0.01084400
## 1384 4.512763      0.0000069043  0.01137800
## 1054 4.505082      0.0000071553  0.01179200
## 149  4.454081      0.0000090574  0.01492700
## 434  4.405401      0.0000113170  0.01865100
## 586  4.360269      0.0000138860  0.02288500
```

Residual autocorrelation

```
library(lmtest)
dwtest(model)
```

```
##
## Durbin-Watson test
##
## data:  model
## DW = 1.2449, p-value = 0.7605
## alternative hypothesis: true autocorrelation is greater than 0
```

There seems to be no evidence of correlation as the p-value is greater than 0.05.

Conclusion

The multiple regression was able to provide a better model for predicting the `Life expectancy`. However, diagnostic plots pointed out a couple of non-normality within the residual. Selection of variables through step-wise selection did not provide a better model to what was earlier at hand based on the significant variables, but had a higher AIC and BIC values. The data requires further analysis and comparison for the individual variables to be able to well assess their predicatbility of the `Life expectancy`.

Appendix

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.936
## [2] LC_CTYPE=Chinese (Simplified)_China.936
## [3] LC_MONETARY=Chinese (Simplified)_China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.936
## system code page: 1252
##
## attached base packages:
## [1] stats      graphics  grDevices utils      datasets  methods   base
##
## other attached packages:
##  [1] lmtest_0.9-38  zoo_1.8-8      broom_0.7.2    plotly_4.9.2.1
##  [5] reshape_0.8.8  forcats_0.5.0  stringr_1.4.0  dplyr_1.0.2
##  [9] purrr_0.3.4    readr_1.4.0    tidyr_1.1.2    tibble_3.0.4
## [13] ggplot2_3.3.2  tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] httr_1.4.2      jsonlite_1.7.1  viridisLite_0.3.0
##  [4] carData_3.0-4   modelr_0.1.8    assertthat_0.2.1
##  [7] highr_0.8       cellranger_1.1.0 yaml_2.2.1
## [10] ggrepel_0.8.2   lattice_0.20-41 pillar_1.4.7
## [13] backports_1.2.0 glue_1.4.2       digest_0.6.27
## [16] rvest_0.3.6     snakecase_0.11.0 colorspace_2.0-0
## [19] htmltools_0.5.0 plyr_1.8.6       infer_0.5.3
## [22] pkgconfig_2.0.3 haven_2.3.1      scales_1.1.1
## [25] openxlsx_4.2.3  rio_0.5.16       generics_0.1.0
## [28] farver_2.0.3    car_3.0-10       ellipsis_0.3.1
## [31] withr_2.3.0     janitor_2.0.1    lazyeval_0.2.2
## [34] formula.tools_1.7.1 cli_2.2.0        magrittr_2.0.1
## [37] crayon_1.3.4    readxl_1.3.1     evaluate_0.14
## [40] fs_1.5.0        fansi_0.4.1      operator.tools_1.6.3
## [43] xml2_1.3.2      foreign_0.8-80   tools_4.0.3
## [46] data.table_1.13.4 hms_0.5.3        lifecycle_0.2.0
## [49] munsell_0.5.0   reprex_0.3.0     zip_2.1.1
## [52] compiler_4.0.3  moderndive_0.5.0 rlang_0.4.8
## [55] grid_4.0.3      rstudioapi_0.13  htmlwidgets_1.5.3
## [58] crosstalk_1.1.0.1 labeling_0.4.2    rmarkdown_2.5
## [61] gtable_0.3.0    abind_1.4-5      DBI_1.1.0
## [64] curl_4.3         R6_2.5.0         lubridate_1.7.9.2
## [67] knitr_1.30       utf8_1.1.4       stringi_1.5.3
## [70] Rcpp_1.0.5       vctrs_0.3.5      dbplyr_2.0.0
## [73] tidyselect_1.1.0 xfun_0.19
```